CDF Note 9850

Measurement of the $t\bar{t}$ + jet Cross Section with 4.1 fb^{-1}

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: July 15, 2009)

The first measurement of the cross section of $t\bar{t}$ associated with an additional hard jet ($t\bar{t}$+jet) has been performed with 4.1 fb^{-1} of collected data at CDF. The measurement is an important test of perturbative QCD, as NLO effects play an important role in the calculation of the theoretical cross section. In addtion, most top events at the LHC will be produced with additional jets, and therefore this process is a substantial background for many new physics signals. The measurement is performed using b-tagged events in the lepton plus jets channel. A data-driven approach is used to predict the background content, and a 2D likelihood is formed to simultaneously measure the $t\bar{t}$+jet and tt without jet cross sections. The measured result is $\sigma_{t\bar{t}+j} = 1.6 \pm 0.2_{stat} \pm 0.5_{syst}$ pb, which is in agreement with the standard model prediction $\sigma_{t\bar{t}+j} = 1.79^{+0.16}_{-0.31}$ pb.

I. INTRODUCTION

Top physics plays an important role both at the Tevatron and at the LHC. The top quark is by far the heaviest elementary fermion in the Standard Model and may have a role in electroweak symmetry breaking. There may be deviations from pointlike behavior for the top quark leading to the presence of anomalous couplings, and in particular to anomalous couplings to the gluon. Such anomalous couplings of the gluon may manifest themselves in deviations from the pQCD predictions for jets accompanying the top-antitop pair.

The $t\bar{t}$ + jet cross section has never been explicitly measured at the Tevatron, although a sizeable fraction of $t\bar{t}$ events at the Tevatron are accompanied by an additional jet . The theoretical cross section has been known to leading order in the QCD coupling constant for many years, and has recently been calculated to next-to-leading order (NLO) [1]. The NLO calculation results in a significant decrease in the theoretical uncertainty; that, in conjunction with the large integrated luminosity accumulated by CDF allows for a precision comparison of data to theory to be performed for the first time. Such a comparison is interesting in its own right, but also as a preview of the LHC, where, due to the production of the top quark pairs by low x gluons, essentially every $t\bar{t}$ event contains an additional jet, and $t\bar{t}$ + jet(s) events form a background to many important signatures for possible new physics [2].

The NLO calculation of the $t\bar{t}$ + jet cross section is perhaps the most difficult pQCD calculation performed to date. There is significant complexity due to the fact that: (1) all of the partons are colored, (2) there is an additional mass scale given by the top quark mass, (3) the infrared structure is complex, (4) there are many diagrams, leading to large expressions and (5) the presence of 1-loop pentagon diagrams. This calculation is also an important component for the calculation of the inclusive $t\bar{t}$ cross section to NNLO. If we consider jet production with a threshold of 20 GeV/c, then there are two very different "reasonable scales that can be used in the calculation: the top mass and the tranverse momentum of the jet. At LO, the disparity in these two scales leads to a large disparity in the size of the predicted cross section. This scale uncertainty is greatly reduced at NLO.

The measurement of the $t\bar{t}$ asymmetry in inclusive events in CDF has resulted in a great deal of interest from the theoretical community, due to the larger than expected value observed. Given the possible importance of this result, it is crucial to understand as many aspects of the $t\bar{t}$ production as possible. The asymmetry for inclusive $t\bar{t}$ production appears only at the 1-loop level, and thus a NLO calculation for the inclusive cross section provides only a LO calculation of the asymmetry. On the other hand, an asymmetry (of the opposite sign as that for the inclusive case, and thus nominally a dilution of this asymmetry), is present at LO for $t\bar{t}$ + jet production; thus the NLO calculation for this cross section is truly a NLO calculation for the $t\bar{t}$ + jet asymmetry. It is somewhat surprising that the 1-loop corrections to this process greatly reduce the size of the asymmetry and is worth investigating experimentally.

II. OVERVIEW OF THE MEASUREMENT

The measurement is performed in the lepton + jets channel, where $t\bar{t} \to W^+W^-b\bar{b}$ and one of the W-bosons decays into a lepton and neutrino and the other to quarks. Candidate events are collected through high p_T lepton triggers and a $\not\!\!E_T$ + Jets trigger. Each event is required to have one high energy electron or muon with $E_T(p_T) > 20$ GeV and at least three jets with $E_T > 20$ GeV and $\eta < 2.0$. One of the jets must be found to have a secondary vertex displaced from primary as evidence of coming from a b-jet from top decay. Because of the presence of a neutrino, we require a large amount of missing transverse energy ($\not\!\!E_T > 20$ GeV). Finally, to further reduce backgrounds the total sum of the transverse energy in the detector is required to be at least 220 GeV ($H_T > 220$ GeV).

Several unwanted processes fake our top signal and contribute as background to the cross section measurement. These include W+jet, QCD, di-boson, and some smaller electroweak processes. The amount of these background contributions are derived from a mixture of Monte Carlo and data-driven techniques, which will be discussed below. A two-dimensional likelihood is formed from the total predicted number of events and the data. The measured cross sections and statistical uncertainty are then extracted for $t\bar{t} + j$ events and $t\bar{t}$ without jets.

III. BACKGROUND ESTIMATE

We take a data-driven approach to backgrounds due to inadequacies in the Monte Carlo to model heavy flavor associated with the production of a W boson, tagging of bottom jets, and difficulties associated with modeling the QCD contribution. The technique is sequential in that each step depends on the previous. The final result is a complete prediction for the process content in the lepton plus jets data sample. In the following we will go step by step through the procedure.

A. Monte Carlo Based Backgrounds

A few of the backgrounds which are considered a small contribution to the overall process content and $t\bar{t}$ (which is an important point as we will discuss later) are calculated based on Monte Carlo efficiencies. Several electroweak processes contribute to the lepton plus jets sample such as WW, WZ, ZZ, and Z $\rightarrow$ *jets* events. They exist in the sample because each process can produce a real lepton and neutrino, as well as a number of jets. The numbers in our sample are estimated using the theoretical cross section, the luminosity of the sample, trigger efficiency, and an overall selection efficiency derived from Monte Carlo simulation of the processes in question. The calculated number in our sample is given by

$$N_{p\bar{p}\rightarrow X} = \sigma_{p\bar{p}\rightarrow X} \cdot A \cdot \int dt \cdot \mathcal{L} \quad (1)$$

$$N^{tag}_{p\bar{p}\rightarrow X} = \sigma_{p\bar{p}\rightarrow X} \cdot A \cdot \epsilon \cdot \int dt \cdot \mathcal{L} \quad (2)$$

where $\sigma_{p\bar{p}\rightarrow X}$ is the theoretical cross section, $\int dt \cdot \mathcal{L}$ is the total luminosity, A is the pre-tagged selection acceptance derived from Monte Carlo, and ϵ is the tagged selection efficiency. As for top, the acceptance and tagging efficiencies are corrected for trigger efficiencies and tagging.

B. Non-W Based Background Estimate

To determine the non-W fraction in both the pretag and tagged sample, we fit the $\not\!\!E_T$ distribution of a non-W template and a MC signal template to data.

Both data and model templates are fitted to the $\not\!\!E_T$ distribution of isolated pretag data events using a binned likelihood fit. Once the fraction is calculated the normalization is simply:

$$N^{pretag}_{QCD} = F_{QCD} \cdot N_{pretag} \quad (3)$$

The same general procedure is performed for the tagged sample.

$$N^{tag}_{QCD} = F_{QCD} \cdot N_{tag} \quad (4)$$

C. W + Heavy Flavor

In the pretag data sample, W plus jets is the dumping ground for all events that are not considered QCD, electroweak, or top. The W plus jets normalization is calculated by subtracting the MC-based processes and the QCD from data as shown in equation 5.

$$N^{pretag}_{W\,+\,Jets} = N_{pretag} \cdot (1 - F^{pretag}_{QCD}) - N^{pretag}_{ewk} - N^{pretag}_{top} \quad (5)$$

For the tagged estimate, the W plus jets sample is broken down into two categories: heavy and light flavor. Each of these processes produces a tagged jet very differently and therefore requires different treatment in calculating the normalization.

The contribution of the heavy flavor background to our signal region is calculated by equation 6.

$$N^{tag}_{W+hf} = (N_{pretag}(1 - F_{QCD}) - N_{EW} - N_{singletop} - N_{t\bar{t}}) \cdot f_{HF} \cdot K \cdot \epsilon \quad (6)$$

The number of events predicted in QCD, Electroweak, singletop, and $t\bar{t}$ is subtracted from the pretag sample, leaving an estimate for the number of events with a W-boson. The fraction of these events with jets matched to heavy flavor quarks, f_{HF}, is calculated from a detailed Monte Carlo simulation Alpgen [5], which includes all possible processes contributing to the production of a single real W-boson. This fraction is corrected by a scale factor which is a correction to the Monte Carlo heavy flavor fraction. The HF correction factor is calculated in the 1 jet bin and applied to the rest of the sample. ϵ is the tagging efficiency. f_{HF} and ϵ are calculated for $Wb\bar{b}$, $Wc\bar{c}$, and Wc separately, which define the rates for each of these processes. Only the heavy flavor fraction relies on Monte Carlo, the absolute normalization is derived from the pretag sample in data. The HF correction is derived by a Neural Network fit to variables sensitive to jets matched to heavy flavor and light flavor.

Process	1jet	2jets	3jets	4jets	5jets
Pretag Events	7445	10947	6380	2724	782
Wbb	50.2 ± 15.5	176.2 ± 54.3	128.4 ± 39.8	50.9 ± 16.9	10.2 ± 6.9
Wcc	24.4 ± 7.7	76.9 ± 24.3	65.8 ± 20.8	27.3 ± 9.2	6.0 ± 4.0
Wc	32.6 ± 10.3	75.2 ± 23.7	41.6 ± 13.2	13.1 ± 4.4	2.4 ± 1.6
Mistags	111.4 ± 11.2	181.7 ± 26.8	101.2 ± 18.2	33.2 ± 9.4	6.2 ± 7.4
Non-W	41.6 ± 12.5	116.4 ± 34.9	71.7 ± 21.5	25.5 ± 20.4	9.3 ± 7.5
WW	2.9 ± 0.3	19.0 ± 2.5	14.8 ± 2.0	6.1 ± 0.8	2.0 ± 0.2
WZ	1.0 ± 0.1	7.1 ± 0.8	5.0 ± 0.6	1.9 ± 0.2	0.5 ± 0.1
ZZ	0.1 ± 0.0	0.9 ± 0.1	1.2 ± 0.2	0.5 ± 0.1	0.2 ± 0.0
Z+jets	3.8 ± 0.4	16.3 ± 1.9	16.7 ± 2.1	6.6 ± 0.8	1.8 ± 0.2
Single Top (s-channel)	1.2 ± 0.1	32.6 ± 3.2	16.5 ± 1.6	4.1 ± 0.4	0.8 ± 0.1
Single Top (t-channel)	0.4 ± 0.0	32.9 ± 2.9	18.7 ± 1.6	4.9 ± 0.4	0.9 ± 0.1
$t\bar{t}$ + 0j (5.5 pb)	8.6 ± 1.7	179.3 ± 35.0	534.4 ± 104.2	555.1 ± 108.1	105.7 ± 20.6
$t\bar{t}$ + j (1.6 pb)	0.5 ± 0.3	16.4 ± 10.3	86.7 ± 54.5	163.1 ± 102.6	182.1 ± 114.5
Total Prediction	278.6 ± 37.2	930.9 ± 117.3	1102.8 ± 144.6	892.3 ± 157.0	328.2 ± 118.1
Observed	304	917	1115	882	329

TABLE I: Background Normalizations for ≥ 1 Tag, ≥ 220 GeV, and $\not{E}_T$ 20 GeV

D. Mistags

A secondary vertex is mistakenly reconstructed when poorly reconstructed tracks seem to cross each other near the origin. A secondary vertex that does not originate from heavy flavor quarks is called a mistag.

The negative tag rate is found to be well parametrized by five jet variables (jet E_t, number of good SVX tracks, sum of all jet E_t in the event, jet η, jet ϕ) and measured in a very high statistics sample derived from triggers on 50 GeV jets. In any subsequent analysis this parametrization then gives the probability that a jet with given values of the tag parametrization variables will be negatively tagged. The negative tag probability of an event is taken to be the sum of the probabilities of all the jets in the event. Studies in large control samples derived from jet triggers with different energy thresholds (20 GeV, 75 GeV, 100 GeV) show good agreement between the prediction and the actual number of negative tags.

This technique is applied to estimate the number of events in our sample due to mistags in W + light flavor events. The predicted number of background events from W + light flavor (W+lf) processes is:

$$N^{tag}_{W+lf} = \frac{N_-}{N_{pre}} \cdot (N_{pre} - N^{t\bar{t}}_{pre} - N^{QCD}_{pre} - N^{W+hf}_{pre} - N^{EW}_{pre} - N^{singletop}_{pre}) \tag{7}$$

The predicted amount of $t\bar{t}$, QCD, W+hf, Electroweak, and single top background events is subtracted from the total pretag sample leaving an estimate for the W+lf fraction. The predicted number of mistagged W+lf events is the W+lf fraction multiplied by the predicted amount of mis-tagged events from the pretag data.

E. Full Background Prediction

The following is the background estimate used in our top cross section measurement utilizing 4.1 fb^{-1} of collected data. Inclusive trigger tables for ≥ 1 Tags are shown in Table I. A histogram representing the predicted number of events and data is shown in Figure 1.

IV. CALCULATING THE CROSS-SECTION

With the background estimate in hand it we now perform a simultaneous measurement of the tt+0j and tt+j cross section. Because the background estimate is dependent on the top cross-section, extracting the measured value is not so simple. Instead, we construct a poisson likelihood where we take into account the background dependence. To extract the measured values we construct a 2D likelihood from the data and prediction for events with three, four, or five jets.

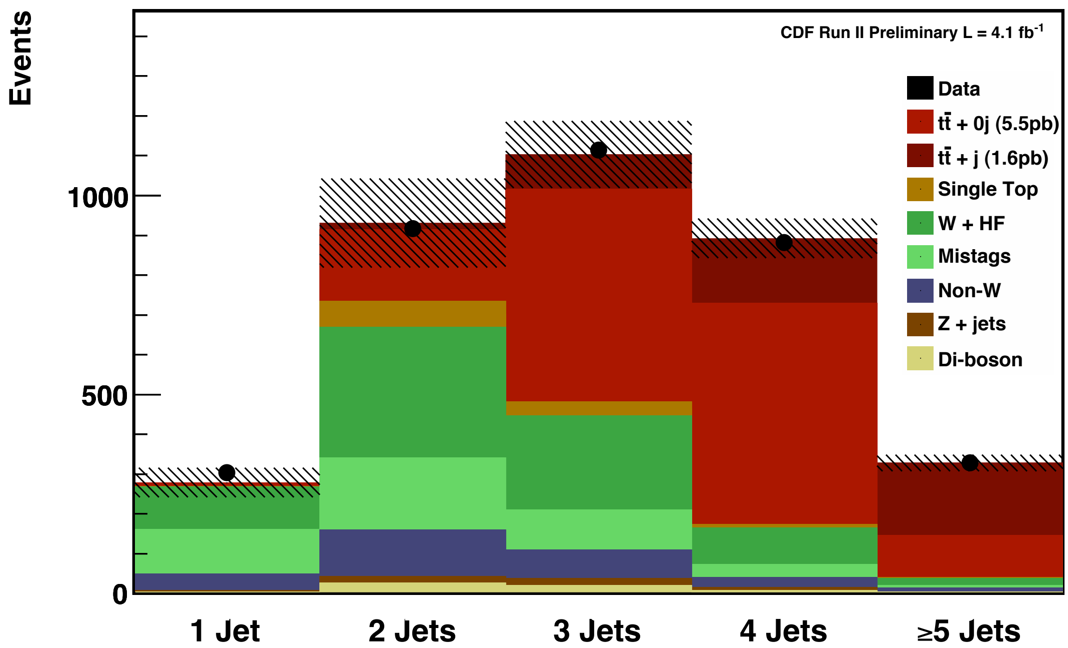

FIG. 1: Predicted vs Observed as a function of jet multiplicity

$$P_i = \frac{\lambda_i^{k_i} \cdot e^{-\lambda_i}}{k_i!} \tag{8}$$

where k is the number of events in data with "i" jets, and λ is the predicted number of events with "i" jets. More specifically:

$$\lambda = A_{0j} \cdot \epsilon_{0j} \cdot \mathcal{L} \cdot \sigma_{t\bar{t}}^{0j} + A_{+j} \cdot \epsilon_{+j} \cdot \mathcal{L} \cdot \sigma_{t\bar{t}}^{+j} + Bkg(\sigma_{t\bar{t}}^{0j}, \sigma_{t\bar{t}}^{+j}) \tag{9}$$

A_x is the acceptance, ϵ is the tagging efficiency, $\mathcal{L}$ is the luminosity, and Bkg is the predicted background. The likelihood is then:

$$L = -ln(P_3 \cdot P_4 \cdot P_5) \tag{10}$$

The likelihood is calculated for several values of the cross-section and the resulting points are fit to a two-dimensional second order polynomial. The minimum of this curve is taken as the measured value. The result for our optimized selection, $H_t \geq 220$ GeV and $\not{E}_T \geq 20$ GeV, is shown in Figure 2. The measured values with statistical uncertainty are:

$$\sigma_{t\bar{t}+0j} = 5.5 \pm 0.4_{stat} \;\; pb \tag{11}$$

$$\sigma_{t\bar{t}+j} = 1.6 \pm 0.2_{stat} \;\; pb \tag{12}$$

Kinematics plots normalized to method II predictions using the measured cross sections are shown in appendix **??**.

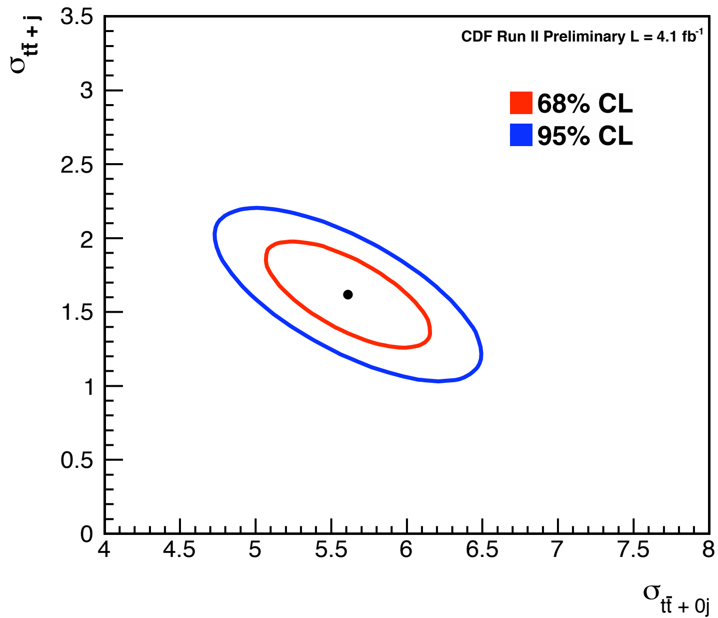

FIG. 2: Likelihood Curve For Measured $t\bar{t}+0j$ and $t\bar{t}+j$ Cross-Section

V. SYSTEMATICS

Systematic uncertainties in our measure result are calculated by varying a given parameter within it's uncertainty and redoing the entire measurement. Each systematic is described below along with any relevant quantities. The individual evaluated systematic uncertainties are shown in Table II at the end of the section.

A. JES

The energy of jets measured by the calorimeters is subject to multiple systematic uncertainties. We study the effect on the measurement by varying the JES for our top signal Monte Carlo and background models and then re-performing the measurement. The effect of JES on this measurement is mainly through the acceptance of signal and background.

B. ISR/FSR

The measured value will be effected if we are over or under estimating the amount of initial or final state radiation present in top events. To study this effect, we replace our standard top Monte Carlo model with top Monte Carlo where the radiation has been increased or decreased and the measurement is redone.

C. Tagging

Because Monte Carlo does not model SecVtx tagging properly, a scale factor is applied to each tagged jet matched to heavy flavor, and the corresponding event then re-weighted. The scale factor is derived from data and has an uncertainty associate with it which leads to a systematic on the measurement. The effect on the measured value is calculated by fluctuating the scale factor within it's uncertainty, applying it to each appropriate jet, calculating the new event weights, and repeating the measurement.

D. Mis-tags

Mistags are so badly modeled in Monte Carlo that we scrap any mis-tagged jet and use a data-based parameterization called the mistag matrix to predict the probability that any given jet is mis-tagged. The mistag rate on any jet is fluctuated within error and the entire measurement is repeated to quantify the effect.

E. QCD Fractions

To estimate the uncertainty on the QCD fraction, fits are performed with different binning and different models. The resulting difference in the fits is 30% which is taken as a systematic uncertainty in the measurement.

F. Heavy Flavor Corrections

The correction to the W + heavy flavor fractions has an uncertainty derived from the Neural Network fits in the 1 and 2 jet bin as well as the fits to bottom and charm separately. A 30% uncertainty is taken on the derived correction to cover the range of fitted values.

G. MC Generator

Differences in Monte Carlo models for parton showering are studied simply by replacing our $t\bar{t}$ pythia model with the other most popular generator, Herwig, and repeating the measurement. Herwig is separated into $t\bar{t}+0j$ and $t\bar{t}+j$ events exactly as pythia is and the measurement was repeated.

H. Trigger Efficiency

Detector specific corrections are applied to the Monte Carlo to more correctly model the relative trigger efficiencies between CEM, CMUP, and CMX events. The corrections are data-derived from Z events and have a small uncertainty associated with them. There are two types of corrections, trigger ID and trigger efficiencies. Each are fluctuated with their uncertainty, separately, and the resulting errors are added in quadrature.

I. PDF

Uncertainty in the parton distribution function are evaluated by a re-weighting scheme at the Monte Carlo Truth level. PDF's are reweighted in our signal Monte Carlo to simulate 46 different PDF parameterizations. The measurement is performed for each different parameterization. The result is shown in Figure II.

J. Luminosity

The uncertainty on our calculated luminosity is unfortunately also our largest systematic, which is derived from detector accuracy and the uncertainty on the theoretical cross section for inelastic $p\bar{p}$ collisions. The uncertainty on the luminosity is 5.8%. The luminosity used in the measurement is fluctuated within this uncertainty and the measurement redone.

Systematic	$\Delta\sigma_{0j}$ pb	$\Delta\sigma_{0j}/\sigma_{0j}$ %	$\Delta\sigma_{+j}$	$\Delta\sigma_{+j}/\sigma_{+j}$ %
JES	0.27	4.9	0.48 pb	30.2
BTag SF	0.25	4.6	0.07	4.6
C Tag SF	0.01	0.2	0.01	0.4
Mistag Matrix	0.01	0.2	0.01	0.6
Heavy Flavor Correction	0.36	6.7	0.06	3.4
Luminosity	0.32	5.6	0.10	6.1
QCD Fraction	0.01	0.2	0.01	0.4
ISF/FSR	0.11	2.1	0.07	3.3
MC Generator	0.19	3.5	0.04	2.3
Trigger Eff	0.03	0.6	0.01	0.6
PDF	0.06	1.0	0.01	1.0
Total	0.65 pb	11.8 %	0.47 pb	36.5 %

TABLE II: Systematic Uncertainties for $t\bar{t}+0j$ and $t\bar{t}+j$

VI. RESULT

The first measured cross section of $t\bar{t}$ in association with a hard jet using 4.1 fb^{-1} of collected data is:

$$\sigma_{t\bar{t}+j} = 1.6 \pm 0.2_{stat} \pm 0.5_{syst} \;\; pb \tag{13}$$

which is in agreement with the Stadard Model prediction $\sigma_{t\bar{t}+j} = 1.79^{+0.16}_{-0.31}$ pb from reference [1]. The measured cross section for $t\bar{t}$ without additional radiation is:

$$\sigma_{t\bar{t}+0j} = 5.5 \pm 0.4_{stat} \pm 0.7_{syst} \;\; pb \tag{14}$$

which when combined with $\sigma_{t\bar{t}+j}$ gives the inclusive cross-section:

$$\sigma_{t\bar{t}} = 7.1 \pm 0.3_{stat} \;\; pb \tag{15}$$

In agreement with the Standard Model prediction at $M_t = 172.5$ GeV of 7.4pb [8].

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community's Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] S. Dittmaier, P.Uwer, and S.Weinzierl, "Hadronic top-quark pair production in association with a hard jet at next-to-leading order QCD: Phenomenological studies for the Tevatron and the LHC", arXiv:0810.0452 [hep-ph].
[2] J.M.Campbell, J.W.Huston and W.J.Stirling, "Hard interactions of quarks and gluons: A primer for LHC physics", arXiv:hep-ph/0611148.
[3] `http://www.pa.msu.edu/~huston/SpartyJet/SpartyJet.html`
[4] T.Affolder et al., Phys Rev D, 64, 032002 (2001)
[5] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP 0307:001,2003, hep-ph/0206293.
[6] T.Sjostrand et al., Comput. Phys. Commun. 135, 238, (2001)
[7] G.Corcella et al., JHEP 01,10 (2001)
[8] Moch, Uwer, hep-ph/08072794 (2008)